

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-142488

(43)Date of publication of application : 25.05.2001

(51)Int.Cl.

G10L 15/28

G10L 15/00

G10L 15/02

H04L 12/56

H04L 29/06

H04M 11/00

(21)Application number : 11-326659

(71)Applicant : OKI ELECTRIC IND CO LTD

(22)Date of filing : 17.11.1999

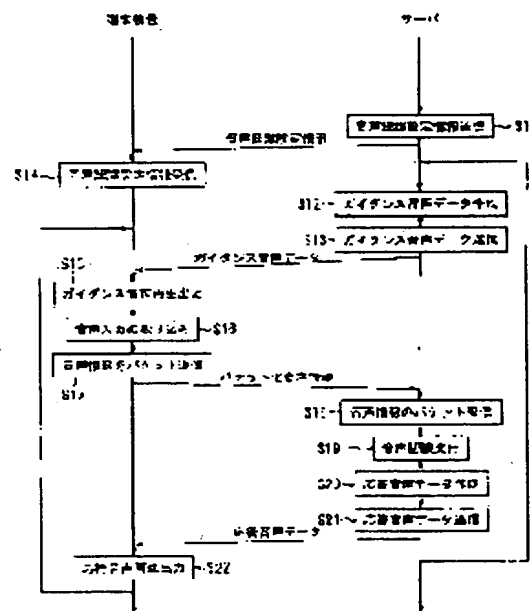
(72)Inventor : WATANABE SATOSHI

(54) VOICE RECOGNITION COMMUNICATION SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a voice recognition communication system which improves the voice recognition performance of voice data between devices on a communication network.

SOLUTION: The voice recognition communication system is provided with a voice input device which takes a sound signal as the input and outputs voice information indicating the sound signal to a network line as a packet, and a voice recognition device which receives the packet transmitted through the network line to perform voice recognition. The voice input device has a transmission means which transmits the packet by one of at least two kinds of communication protocols different by degrees of reliability, and the voice recognition device has a reception means which receives the transmitted packet in accordance with one communication protocol.



LEGAL STATUS

[Date of request for examination]

26.12.2002

[Date of sending the examiner's decision of rejection]

04.04.2005

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-142488

(P2001-142488A)

(43) 公開日 平成13年5月25日 (2001.5.25)

(51) Int.Cl. ⁷	識別記号	F I	テーマコード* (参考)	
G 1 0 L 15/28		H 0 4 M 11/00	3 0 2	5 D 0 1 5
15/00		G 1 0 L 3/00	5 7 1 A	5 K 0 3 0
15/02			5 5 1 A	5 K 0 3 4
H 0 4 L 12/56		5/06	A	5 K 1 0 1
29/06		H 0 4 L 11/20	1 0 2 A	
審査請求 未請求 請求項の数 8 O L (全 10 頁) 最終頁に続く				

(21) 出願番号 特願平11-326659

(22) 出願日 平成11年11月17日 (1999. 11. 17)

(71) 出願人 000000295

沖電気工業株式会社

東京都港区虎ノ門1丁目7番12号

(72) 発明者 渡辺 聡

東京都港区虎ノ門1丁目7番12号 沖電気工業株式会社内

(74) 代理人 100079119

弁理士 藤村 元彦

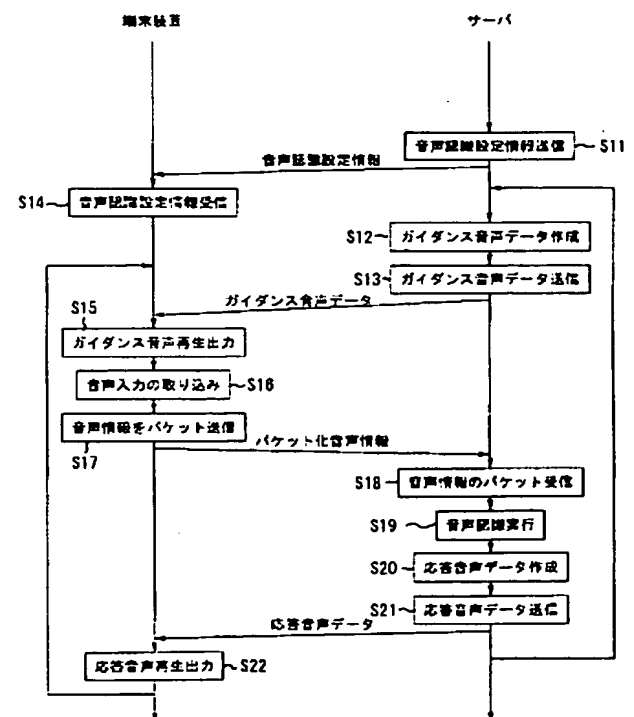
最終頁に続く

(54) 【発明の名称】 音声認識通信システム

(57) 【要約】

【課題】 通信ネットワーク上の装置間における音声データ音声認識性能を向上させることができる音声認識通信システムを提供する。

【解決手段】 音声信号を入力しその音声信号を示す音声情報をパケットとしてネットワーク回線に出力する音声入力装置と、ネットワーク回線を介して伝送されたパケットを受信して音声認識を行う音声認識装置とを備えた音声認識通信システムであって、音声入力装置は、互いに信頼性の程度が異なる少なくとも2種類の通信プロトコルのうちのいずれか1のプロトコルにてパケットを送信する送信手段を有し、音声認識装置は、伝送されてきたパケットを1の通信プロトコルに従って受信する受信手段を有する。



【特許請求の範囲】

【請求項1】 音声信号を入力しその音声信号を示す音声情報をパケットとしてネットワーク回線に出力する音声入力装置と、前記ネットワーク回線を介して伝送されたパケットを受信して音声認識を行う音声認識装置とを備えた音声認識通信システムであって、

前記音声入力装置は、互いに信頼性の程度が異なる少なくとも2種類の通信プロトコルのうちのいずれか1の通信プロトコルにてパケットを送信する送信手段を有し、前記音声認識装置は、伝送されてきたパケットを前記1の通信プロトコルに従って受信する受信手段を有することを特徴とする音声認識通信システム。

【請求項2】 前記音声情報は、音声信号を高圧縮率符号化したデータであることを特徴とする請求項1記載の音声認識通信システム。

【請求項3】 前記音声情報は、音声信号のPCMデータであることを特徴とする請求項1記載の音声認識通信システム。

【請求項4】 前記音声情報は、音声信号の特徴を示すパターンデータであることを特徴とする請求項1記載の音声認識通信システム。

【請求項5】 前記ネットワーク回線におけるパケット伝送はインターネットプロトコルを用いて行うことを特徴とする請求項1記載の音声認識通信システム。

【請求項6】 前記少なくとも2種類の通信プロトコルのうちの一方の通信プロトコルはコネクションレス型の通信プロトコルであり、他方の通信プロトコルは前記コネクションレス型の通信プロトコルより信頼性が高いコネクション型の通信プロトコルであることを特徴とする請求項1記載の音声認識通信システム。

【請求項7】 前記一方の通信プロトコルはUDPであり、前記他方の通信プロトコルはTCPであることを特徴とする請求項6記載の音声認識通信システム。

【請求項8】 前記一方の通信プロトコルはUDPであり、前記他方の通信プロトコルはFTPであることを特徴とする請求項6記載の音声認識通信システム。

【発明の詳細な説明】

【0001】

【発明が属する技術分野】本発明は、音声入力装置において入力された音声について音声入力装置とネットワーク回線を介して接続された音声認識装置において音声認識する音声認識通信システムに関する。

【0002】

【従来の技術】ネットワークを介して双方向の通話を行うシステムとしては、電話回線網及びその電話端末からなる音声通信システムが従来からよく知られている。このような音声通信システムにおいて音声が入力された電話端末から送られてくる音声信号に対して音声認識を行うことがある。かかる音声認識においては、多様な電話端末及び回線から送られてくる多様な品質の音声信号に

対して音声認識を行う必要がある。一般には、回線毎に音声認識のパターンマッチングのために標準パターンや汎用的なパターンが音声認識側に用意されている。受信した音声信号を分析して音声パターンを作成し、それを用意されたパターンと比較することにより、電話端末で入力された音声の内容が判定される。

【0003】また、近年、インターネットを用いて通話を行ういわゆるインターネット電話に代表される音声IP（インターネットプロトコル）通信システムがある。この音声IP通信システムにおいては、音声認識を適用した場合には端末装置で入力された音声を音声認識サーバで音声認識することが行われる。ネットワーク回線では通常、音声データはパケット化して伝送されるので、サーバでは受信したパケットから音声データを取り出して音声認識を行う必要がある。

【0004】

【発明が解決しようとする課題】しかしながら、音声IP通信システムにおいては、ネットワーク上の通信が混雑した状態である場合にはパケットが正常に送り先の音声認識サーバや端末装置に到達しないパケットロス（パケット落ち）が発生することがある。パケットロスが発生した場合に、失われたパケットに相当する部分の情報を得ることができないので、受信側では音声データを全てに亘って復号できないことになる。よって、前後のフレームから予測した音声データで代用することが行われるので、その結果として音声認識性能は低下することになるという問題点があった。

【0005】また、通常、音声IP通信では、ネットワーク帯域の有効利用と遅延防止等の理由から4～8kbps程度の高圧縮率音声符号化を行うために復号しても元の音声情報が変形或いは欠落してしまうので、高精度の音声認識が難しいという問題点もあった。そこで、本発明の目的は、通信ネットワーク上の装置間における音声データ音声認識性能を向上させることができる音声認識通信システムを提供することである。

【0006】

【課題を解決するための手段】本発明の音声認識通信システムは、音声信号を入力しその音声信号を示す音声情報をパケットとしてネットワーク回線に出力する音声入力装置と、ネットワーク回線を介して伝送されたパケットを受信して音声認識を行う音声認識装置とを備えた音声認識通信システムであって、音声入力装置は、互いに信頼性の程度が異なる少なくとも2種類の通信プロトコルのうちのいずれか1のプロトコルにてパケットを送信する送信手段を有し、音声認識装置は、伝送されてきたパケットを1の通信プロトコルに従って受信する受信手段を有することを特徴としている。この構成によれば、通信プロトコルとして少なくとも2種類の通信プロトコルを使用し、パケットロスの発生等の通信状態の悪化に応じて信頼性の高い通信プロトコルを選択的に使用する

ことができる。

【0007】

【発明の実施の形態】以下、本発明の実施例を図面を参照しつつ詳細に説明する。図1は本発明によるシステムの概略構成を示している。このシステムにおいては、端末装置1とサーバ2とがインターネット回線網3を介して接続された構成となっている。

【0008】端末装置1は、図2に示すように、サーバ2との接続処理を行う接続制御部11と、サーバ2からインターネット回線網3を介して供給された音声データをアナログの音声信号としてスピーカ15に出力する音声出力部12と、マイクロホン16で集音された音声及び音声情報をデータとしてサーバ2に供給するためにインターネット回線網3に送り出す音声入力部13と、端末装置1全体の動作を制御する端末制御部14とを備えている。

【0009】音声出力部12は、インターネット回線網3を介して送られてきた音声データを受信する音声受信部31と、音声受信部31で受信された音声データを復号する音声復号化部32と、復号された音声データをアナログの音声信号に変換してスピーカ15に供給する外部出力部33とからなる。音声入力部13は、マイクロホン16から出力された音声信号をデジタル音声データであるPCM(Pulse Code Modulation)データに変換する外部入力部35と、外部入力部35から供給された音声データを高圧縮率で符号化する音声符号化部36と、インターネット回線網3に接続され、符号化された音声データをUDP(User Datagram Protocol)パケットにてサーバ2に対して送信するUDP音声送信部37と、マイクロホン16への入力音声の特徴を示す認識パラメータを計算して音声パターンを出力する認識パラメータ計算部38と、音声パターンをサーバ2に対して送信する音声パターン送信部39とからなる。

【0010】サーバ2は、図3に示すように、端末装置1との接続処理を行う接続制御部41と、端末装置1からUDPパケットにて送られてきた音声データを受信するUDP音声受信部42と、受信した音声データを復号する音声復号化部43と、復号された音声データの内容を認識して認識結果に対応した音声データを作成する音声対話部44と、音声対話部44から出力された音声データを符号化する音声符号化部45と、インターネット回線網3に接続され、音声復号部45で符号化された音声データを端末装置1に対して送信する音声送信部46と、端末装置1から送られてきた音声パターンデータを受信する音声パターン受信部47と、サーバ2全体を制御するサーバ制御部48とを備えている。

【0011】音声対話部44は、音声対話部44から出力された音声データを分析することにより認識パラメータを計算し、その計算結果として音声パターンを出力する認識パラメータ計算部51と、音声パターンと標準パ

ターンとのマッチング(類似度)を計算して音声データの内容を特定するパターンマッチング部52と、パターンマッチング部52の出力に応じて応答音声データの作成を指示する音声対話制御部53と、音声データを作成する音声作成部54とからなる。

【0012】かかる構成のシステムにおいては、端末装置1は図4に示すように、サーバ2に対して音声通信のための接続要求を行ってサーバ2とのインターネット回線網3を介した接続を確立する(ステップS1)。この接続のためにIP音声通信で規定されている方法、例えば、ITU-T H.323に定められたゲートキーパを介する方法が用いられる。端末装置1とサーバ2との音声通信が確立すると、端末装置1は後述するシーケンス(手順)でサーバ2との音声対話処理を行う(ステップS2)。音声対話処理が終了すると、端末装置1はサーバ2に対して切断要求を行ってサーバ2との接続を切断する(ステップS3)。この切断のためにはIP音声通信で規定されている方法、例えば、ITU-T H.323に定められたゲートキーパを介する方法が用いられる。

【0013】次に、上記のステップS2の音声対話処理のシーケンスについて図5を用いて説明する。サーバ2は、音声認識の設定情報を端末装置1に送信し(ステップS11)、端末装置1のユーザの発声を促すガイダンス音声データを作成する(ステップS12)。このガイダンス音声データは例えば、「発声してください。」の如き発声音を示すデータである。サーバ2は作成したガイダンス音声データを端末装置1に送信する(ステップS13)。すなわち、サーバ制御部48はステップS11では接続制御部41に音声認識の設定情報を端末装置1に送信させる。音声認識の設定情報とは音声認識のための音声データ形式であり、例えば、分析フレーム長: 16msec、分析フレーム周期: 8msec、LPG分析次数: 12次、認識パラメータ: PARCOR係数である。ステップS12ではサーバ制御部48は音声対話制御部53を介して音声作成部54にガイダンス音声データの作成を指示し、音声作成部54において作成されたガイダンス音声データは音声符号化部45で符号化された後、音声送信部46から端末装置1宛のUDP(User Datagram Protocol)パケットにてインターネット回線網3に出力される。

【0014】端末装置1は、音声認識の設定情報を受信してその設定情報を保存し(ステップS14)、またガイダンス音声データを受信してその音声データによる音声再生を行う(ステップS15)。ステップS14では受信された設定情報は接続制御部11を介して端末制御部14に供給されて内部メモリ(図示せず)に保持され、送受信の音声データの形式に反映される。すなわち、その情報に基づいて音声出力部12及び音声入力部13が制御される。ステップS15ではガイダンス音声データは音声受信部31にて受信され、音声復号化部3

2で復号された後、外部出力部33でアナログ音声信号とされる。その音声信号はスピーカ15を駆動し、それにより発声を促す音声再生音がスピーカ15から出力される。

【0015】外部出力部33から音声信号がスピーカ15に出力されると、マイクロホン16によってユーザの発声音が集音され、それが音声信号としてマイクロホン16から音声入力部13に供給される。端末装置1の端末制御部14は音声入力部13を制御することにより、その供給された音声信号が受け入れられてデジタル音声データに変換され（ステップS16）、サーバ2に対して音声情報がパケット化されて送信される（ステップS17）。すなわち、マイクロホン16からの音声信号は外部入力部35にて8kHz/16ビットにデジタル化され、30msec（240サンプル）毎に音声データは音声符号化部36に供給される。音声符号化部36は30msec毎に音声符号化を行い、ビットストリームである符号化音声データを音声送信部37に供給する。この音声符号化としては、例えば、ITU-TG.723.1に規定された方法が用いられる。符号化音声データが供給された音声送信部37は音声データをサーバ2宛のUDP (User Datagram Protocol) パケットにてインターネット回線網3に出力する。UDPは、IETF RFC768で規定されるインターネット上のコネクションレス型トランスポートプロトコルであり、送達確認や順序制御等の機能を備えることなく高速転送を可能にしている。音声送信部37はUDPパケットを作成する際にタイムスタンプやパケット番号を例えば、RTP (Real-time Transfer Protocol) 等の手続きに従って組み込む。RTPは、IETF RFC1889, 1890で規定されているリアルタイム通信用プロトコルである。

【0016】サーバ2は、インターネット回線網3を介して伝送されて来た音声情報を含むパケットを受信すると（ステップS18）、音声認識を行い（ステップS19）、その音声認識結果に応じて応答音声データを作成し（ステップS20）、作成した応答音声データを端末装置1に対して送信する（ステップS21）。具体的に説明すると、インターネット回線網3からUDPパケットを音声受信部42は、受信パケットのタイムスタンプやパケット番号のチェックを例えば、RTP等をプロトコルを使用して行い、異常がなければ、その後、UDPパケットからビットストリームの音声データを抽出して音声復号化部43に供給する。また音声受信部42は、チェックの段階でパケットが完全には受信されていないと判別した場合には全てのビットを0とした仮のビットストリームを音声復号化部43に供給する。音声復号化部43はビットストリームを復号化して8kHz/16ビットの音声データを得て、それをD/A変換することなく認識パラメータ計算部51に供給する。認識パラメータ計算部51は、順次供給される音声データを保持

し、16msec分の音声データを8msec周期で分析し、分析結果として音声データの特徴を示す音声パターンを得る。分析方法としては、例えば、12次のLPC分析等の方法が用いられる。得られた音声パターンはパターンマッチング部52に供給される。パターンマッチング部52は、音声対話制御部53が指定した標準パターンに対して、音声パターン計算部51から供給された音声パターンを照合してマッチング計算を行ってその計算結果を認識結果として音声対話制御部53に供給する。

【0017】音声対話制御部53は認識結果に対応する応答音声（例えば、「東京のお天気は？」）の作成を音声作成部54に指令する。音声作成部54は指令された応答音声に対応する応答音声データを作成して音声符号化部45に供給する。応答音声データは音声符号化部45で符号化された後、音声送信部46から音声送信部46から端末装置1宛のUDPパケットにてインターネット回線網3に出力される。

【0018】端末装置1は応答音声データを受信してその音声データによる音声再生を行う（ステップS22）。このステップS22では応答音声データは音声受信部31にて受信され、音声復号化部32で復号された後、外部出力部33でアナログ音声信号とされる。その音声信号はスピーカ15を駆動し、それによりサーバ2からの応答音声スピーカ15から出力される。

【0019】その後、端末装置1はステップS15に戻って上記の動作を繰り返し、またサーバ2はステップS22の実行後、ステップS12に戻って上記の動作を繰り返す。上記のステップS16～S22の動作については、次のように行うこともできる。なお、ステップS16～S22の2つの動作のうちからいずれか一方の動作がサーバ制御部48の指示（ステップS11の音声認識の設定情報）に応じて選択されて実行される。

【0020】端末装置1においては、マイクロホン16からのユーザの発声音を示す音声信号が音声入力部13に供給されると、そこで8kHz/16ビットにデジタル化され、30msec（240サンプル）毎に音声データは認識パラメータ計算部38に供給される。認識パラメータ計算部38は順次供給される音声データを保持し、16msec分の音声データを8msec周期で分析し、分析結果として音声パターンを得る。分析方法としては、例えば、12次のLPC分析等の方法が用いられる。得られた音声パターンは音声パターン送信部39に供給される。音声パターン送信部39は音声パターンをサーバ2宛のTCP (Transmission Control Protocol) パケットにてインターネット回線網3に出力する。TCPは、IETF RFC793で規定されるインターネット上のコネクションオリエンティッドなトランスポートプロトコルであり、エラー検出及び回復機能等の機能を備え、UDPよりも信頼性の高いデータ通信を行うことができる

プロトコルである。

【0021】サーバ2においては、インターネット回線網3からTCPパケットを認識パラメータ受信部47が受信し、受信したTCPパケットから音声パターンを抽出してパターンマッチング部52に供給する。パターンマッチング部52は、音声対話制御部53が指定した標準パターンに対して、認識パラメータ受信部47から供給された音声パターンを照合してマッチング計算を行ってその計算結果を認識結果として音声対話制御部53に供給する。

【0022】音声IP通信においては、通常使用される圧縮方式のビットレートは、5.3/6.3 kbps (ITU-T G.723.1)や8 kbps (ITU-T G.729)である。一方、この実施例の音声パターンのビットレートは分析フレーム周期8 msec、LPG分析次数12次の場合には、48 kbpsの如く高い。しかしながら、単語発声では1語当たり1秒程度と短時間であり、再送機能を備えたTCPを用いた場合の遅延は一般に許容範囲内である。

【0023】かかる実施例においては、UDPパケットと比べて信頼性が高いTCPパケットを用いて音声パターンを伝送するので、パケットロスが発生した区間の音声パターンが欠落する可能性が低くなり、音声認識の精度を向上させることができる。また、かかる実施例においては、符号化/復号化を行うことなく、マイクロホンに入力された音声から音声パターンを直接抽出しているので、符号化の際の圧縮によって音声パターンが変形するという影響を受けることがなく、また各圧縮方法の違いに依存することなく、音声パターンを得ることができる。よって、音声認識の精度を向上させることができる。

【0024】更に、かかる実施例においては、認識パラメータの計算を端末装置で行うので、サーバの構成を簡略化できると共に、サーバにおける認識パラメータ計算に要する処理量を削減することができる。サーバの処理量を減らすことはサーバが多数の端末装置に対して同時に応答するような場合に特に有効である。なお、かかる実施例においては、プロトコルとしてTCPを用いて音声パターンの送受信が行われるが、ユーザ発声時間分の遅延を考慮して音声パターンをファイル化すれば、FTP (File Transfer Protocol)等のアプリケーションプロトコルを用いても良い。

【0025】更に、かかる実施例においては、サーバ制御部48が音声情報の伝送のための通信プロトコルをコネクションレス型プロトコルのUDP及びコネクション型プロトコルのTCPのうちのいずれか一方を端末装置1に指示するが、その指示ではUDPが優先され、サーバ2で受信したUDPパケットのパケットロスの単位時間当たりの発生率が所定値より大となるとTCPへの通信プロトコルの切替を端末装置1に指示しても良い。

【0026】また、かかる実施例においては、認識パラメータ計算部38を端末装置1内にいわゆるハード的に備えているが、ネットワークを介してダウンロードされたソフトウェアの実行によって認識パラメータ計算部を形成しても良い。図6及び図7は本発明の他の実施例として図1の端末装置1及びサーバ2の内部構成を示している。図6に示した端末装置1においては、PCMデータ送信部40が備えられている。PCMデータ送信部40の入力は外部入力部35に接続され、出力はインターネット回線網3に接続されている。このPCMデータ送信部40は図2に示した端末装置1内の認識パラメータ計算部38及び音声パターン送信部39に代わって備えられている。その他の構成は図2に示した端末装置1と同様である。

【0027】図7に示したサーバ2においては、PCMデータ受信部50が備えられている。PCMデータ受信部50の入力はインターネット回線網3に接続され、出力は認識パラメータ計算部51に接続されている。このPCMデータ受信部50は図3に示したサーバ2内の音声パターン受信部47に代わって備えられている。その他の構成は図3に示したサーバ2と同様である。

【0028】かかる図6及び図7に示した端末装置1及びサーバ2を備えた構成のシステムにおいても、上記の図4に示したように、音声通信の接続確立(ステップS1)、音声対話処理(ステップS2)及び音声通信の接続切断(ステップS3)の如き動作が行われる。上記のステップS2の音声対話処理のシーケンスについては図5に示した如き動作が行われる。

【0029】図6及び図7に示した端末装置1及びサーバ2を備えた場合には、図5のステップS16～S22の動作については、次のように行うこともできる。端末装置1においては、マイクロホン16からのユーザの発声を示す音声信号が音声入力部13に供給されると、そこで8 kHz/16ビットにディジタル化され、30 msec (240サンプル)毎に音声データはPCMデータ送信部40に供給される。PCMデータ送信部40は順次供給されるPCMデータを保持し、その音声データをサーバ2宛のTCPパケットにてインターネット回線網3に出力する。TCPは、上記したように、IETF RFC793で規定されるインターネット上のコネクションオリエンティッドなトランスポートプロトコルであり、エラー検出及び回復機能等の機能を備えた信頼性の高いデータ通信を行うことができるプロトコルである。

【0030】サーバ2においては、インターネット回線網3からTCPパケットをPCMデータ受信部50が受信し、受信したTCPパケットからPCM音声データを抽出して認識パラメータ計算部51に供給する。認識パラメータ計算部51は、順次供給される音声データを保持し、16 msec分の音声データを8 msec周期で分析し、分析結果として音声データの特徴を示す音声パ

ーンを得る。分析方法としては、例えば、12次のLPC分析等の方法が用いられる。得られた音声パターンはパターンマッチング部52に供給される。パターンマッチング部52は、音声対話制御部53が指定した標準パターンに対して、認識パラメータ受信部47から供給された音声パターンを照合してマッチング計算を行ってその計算結果を認識結果として音声対話制御部53に供給する。

【0031】なお、音声IP通信において、通常使用される圧縮方式のビットレートは、5.3/6.3kbps (ITU-T G.723.1)や8kbps (ITU-T G.729)である。一方、この実施例のPCMデータのビットレートは128kbpsの如く高い。しかしながら、単語発声では1語当たり1秒程度と短時間であり、再送機能を備えたTCPを用いた場合の遅延は一般に許容範囲内である。

【0032】かかる実施例においては、UDPパケットと比べて信頼性が高いTCPパケットを用いてPCM音声データを伝送するので、パケットロスが発生した区間の音声データが欠落する可能性が低くなり、音声認識の精度を向上させることができる。また、かかる実施例においては、符号化/復号化を行うことなく、マイクロホンに入力された音声から音声パターンを直接抽出しているので、符号化の際の圧縮によって音声パターンが変形するという影響を受けることがなく、また各圧縮方法の違いに依存することなく、音声パターンを得ることができる。よって、音声認識の精度を向上させることができる。

【0033】更に、かかる実施例においては、認識対象となるPCMデータをサーバ2では符号化歪みが全く無い状態で得ることができるので、異なる分析条件で再認識を行うことができ、サーバでの認識条件の自由度が増加する。また、かかる実施例においては、端末装置1にPCMデータ送信部40を加えただけの構成で音声認識性能の向上を図ることができる。これはLAN用電話機等の処理能力の低い端末装置でも容易に実現が可能な構成であり、音声認識性能の向上を図ることができる。

【0034】なお、かかる実施例においては、プロトコルとしてTCPを用いてPCMデータの送受信が行われるが、ユーザ発声時間分の遅延を考慮してPCMデータをファイル化すれば、FTP等のアプリケーションプロトコルを用いても良い。図8及び図9は、本発明の他の実施例として図1の端末装置1及びサーバ2の内部構成を示している。図8に示した端末装置1においては、符号化された音声データをTCPパケットにてサーバ2に対して送信するTCP音声送信部55が備えられている。TCP音声送信部55の入力は音声符号化部36に接続され、出力はインターネット回線網3に接続されている。その他の構成は、図2に示した端末装置1内の認識パラメータ計算部38及び音声パターン送信部39が備えられていないことを除いて図2に示した端末装置1

と同様である。

【0035】図9に示したサーバ2においては、端末装置1からTCPパケットにて送られてきた音声データを受信するTCP音声受信部56が備えられている。TCP音声受信部56の入力はインターネット回線網3に接続され、出力は音声復号化部43の入力に接続されている。その他の構成は図3に示したサーバ2内の音声パターン受信部47が備えられていないことを除いて図3に示したサーバ2と同様である。

【0036】かかる図8及び図9に示した端末装置1及びサーバ2を備えた構成のシステムにおいても、上記の図4に示したように、音声通信の接続確立(ステップS1)、音声対話処理(ステップS2)及び音声通信の接続切断(ステップS3)の如き動作が行われる。上記のステップS2の音声対話処理のシーケンスについては図5に示した如き動作が行われる。

【0037】図8及び図9に示した端末装置1及びサーバ2を備えた場合には、図5のステップS16～S22の動作については、次のように行うこともできる。端末装置1においては、マイクロホン16からのユーザの発声を示す音声信号が音声入力部13に供給されると、そこで8kHz/16ビットにデジタル化され、30msec(240サンプル)毎に音声データは音声符号化部43に供給される。音声符号化部36は30msec毎に音声符号化を行い、ビットストリームである符号化音声データをTCP音声送信部55に供給する。音声送信部55は順次供給される符号化音声データを保持し、その音声データをサーバ2宛のTCPパケットにてインターネット回線網3に出力する。

【0038】サーバ2においては、インターネット回線網3からTCPパケットを音声受信部56は、受信パケットのタイムスタンプやパケット番号のチェックを例えば、RTP等をプロトコルを使用して行い、異常がなければ、その後、TCPパケットからビットストリームの音声データを抽出して音声復号化部43に供給する。音声復号化部43はビットストリームを復号化して8kHz/16ビットの音声データを得て、それをD/A変換することなく認識パラメータ計算部51に供給する。認識パラメータ計算部51は、順次供給される音声データを保持し、16msec分の音声データを8msec周期で分析し、分析結果として音声データの特徴を示す音声パターンを得る。分析方法としては、例えば、12次のLPC分析等の方法が用いられる。得られた音声パターンはパターンマッチング部52に供給される。パターンマッチング部52は、音声対話制御部53が指定した標準パターンに対して、認識パラメータ計算部51から供給された音声パターンを照合してマッチング計算を行ってその計算結果を認識結果として音声対話制御部53に供給する。

【0039】かかる実施例においては、UDPパケット

と比べて信頼性が高いTCPパケットを用いて符号化音声データを伝送するので、パケットロスが発生した区間の音声データが欠落する可能性が低くなり、音声認識の精度を向上させることができる。また、かかる実施例においては、端末装置1にTCP音声送信部40を加えただけの構成で音声認識性能の向上を図ることができる。これはLAN用電話機等の処理能力の低い端末装置でも容易に実現が可能な構成であり、音声認識性能の向上を図ることができる。

【0040】更に、かかる実施例においては、TCPパケットを用いた場合にはUDPパケットを用いた場合と比べて同一の音声情報を送信することになり、通信量の違いはプロトコルの違いに依存するだけの最小限で済むので、最小限の通信量の増加でパケットロスに対処することができる。なお、かかる実施例においては、プロトコルとしてTCPを用いて符号化音声データの送受信が行われるが、ユーザ発声時間分の遅延を考慮して符号化音声データをファイル化すれば、FTP等のアプリケーションプロトコルを用いても良い。また、3種類以上の通信プロトコルから1の通信プロトコルを選択的に用いても良い。

【0041】上記した各実施例においては、認識パラメータをPARCOR係数に限定したが、認識パラメータとしてはLSP係数やLPCケプストラム係数を用いても良い。また、分析フレーム長や分析フレーム周期、分析次数等の分析パラメータやサンプリング周波数、量子化ビット数、バッファサイズ等の各種パラメータは適宜設定することができる。

【0042】また、上記した各実施例においては、端末装置1及びサーバ2を各々1台備えているが、複数の端末装置及び複数のサーバがネットワークにて互いに接続されたシステムにも本発明を適用することができる。更に、上記した各実施例においては、ステップS16～S22の2つの動作、すなわちUDPパケット伝送を用いた音声認識動作とTCPパケット伝送を用いた音声認識動作とのうちからいずれか一方の動作がサーバ制御部48の指示に応じて選択されて実行されるが、端末装置1の端末制御部14の指示に応じて選択されて実行されても良い。

【0043】また、上記した各実施例においては、端末装置1とサーバ2とがインターネット回線網3を介して接続されるが、端末装置1とサーバ2との間を接続する

ネットワーク回線としては公衆電話回線や他の専用回線であっても良い。更に、上記した各実施例においては、音声入力装置である端末装置1にて音声が入力され、音声認識装置であるサーバ2にてその入力音声の音声識別が行われるが、サーバ2に音声入力装置としての機能を備え、端末装置1に音声認識装置としての機能を備えるようにしても良い。

【0044】

【発明の効果】以上の如く、本発明の音声認識通信システムにおいては、通信プロトコルとして少なくとも2種類の通信プロトコルを使用し、パケットロスの発生等の通信状態の悪化に応じて信頼性の高い通信プロトコルを選択的に使用することができるので、通信ネットワーク上の装置間における音声データ音声認識性能を向上させることができる。

【図面の簡単な説明】

【図1】本発明による音声認識通信システムの概略構成を示すブロック図である。

【図2】端末装置の内部構成を示すブロック図である。

【図3】サーバの内部構成を示すブロック図である。

【図4】端末装置とサーバとの間の概略的動作を示すフローチャートである。

【図5】音声対話処理を具体的に示すフローチャートである。

【図6】本発明の他の実施例として端末装置の内部構成を示すブロック図である。

【図7】図6の端末装置に対応したサーバの内部構成を示すブロック図である。

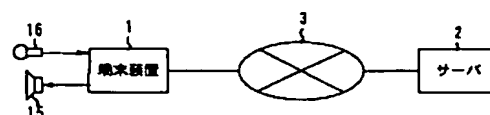
【図8】本発明の他の実施例として端末装置の内部構成を示すブロック図である。

【図9】図8の端末装置に対応したサーバの内部構成を示すブロック図である。

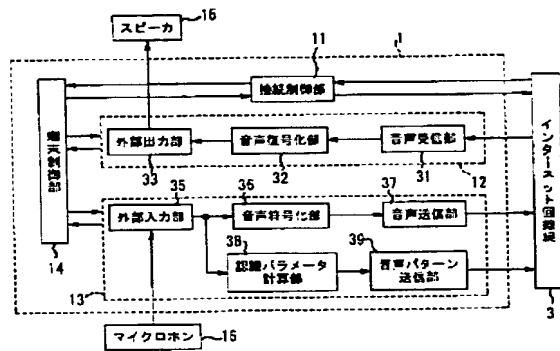
【符号の説明】

- 1 端末装置
- 2 サーバ
- 3 インターネット回線網
- 12 音声出力部
- 13 音声入力部
- 15 スピーカ
- 16 マイクロホン
- 44 音声対話部

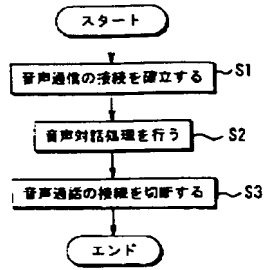
【図1】



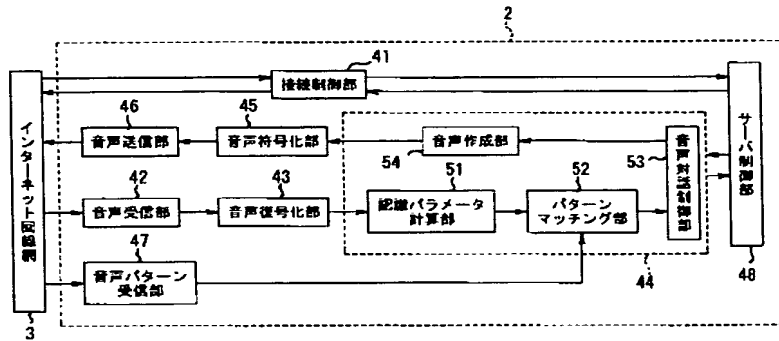
【図2】



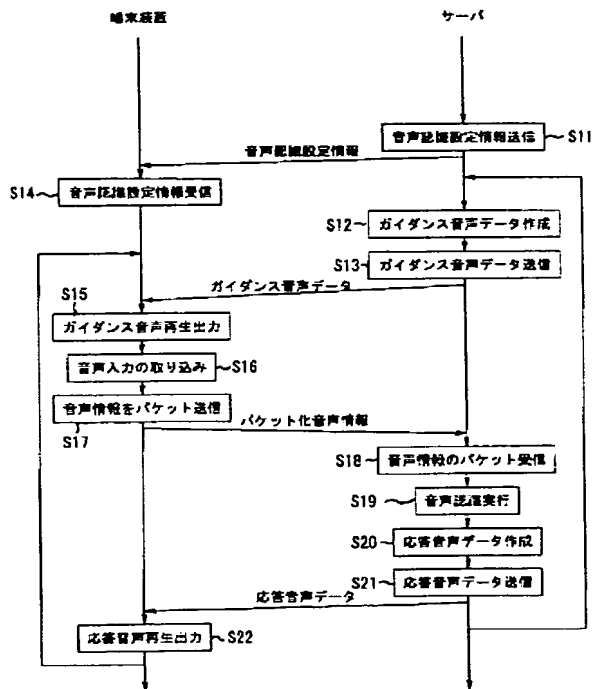
【図4】



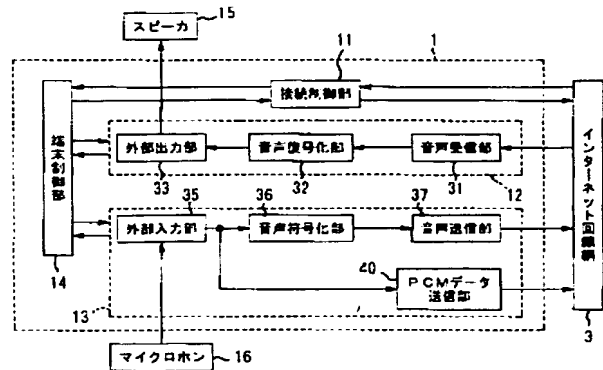
【図3】



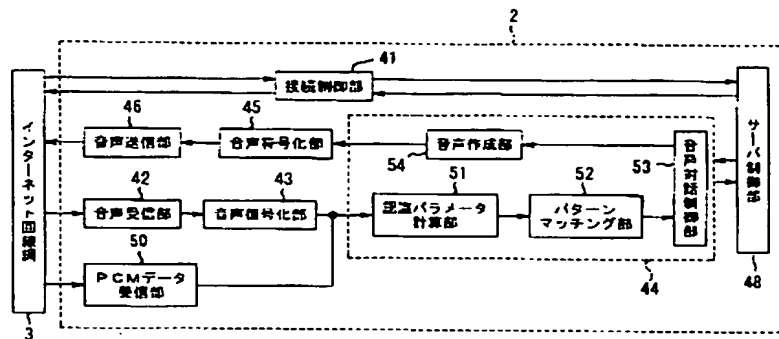
【図5】



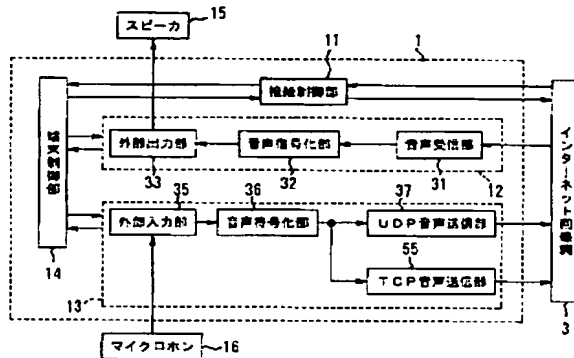
【図6】



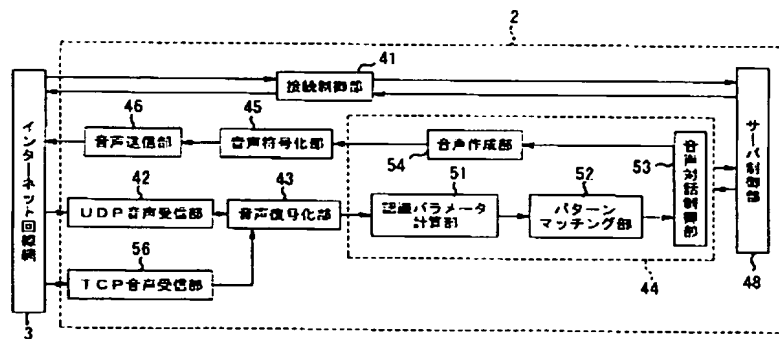
【図 7】



【図 8】



【図 9】



フロントページの続き

(51) Int. Cl. 7

H 0 4 M 11/00

識別記号

3 0 2

F I

H 0 4 L 13/00

テーマコード(参考)

3 0 5 C

F ターム(参考) 5D015 KK02
5K030 GA11 GA12 HA08 HB01 HB18
KA20 LA08 LB18 LB19 MA04
MB04 MB09
5K034 AA05 AA06 CC05 DD01 EE11
FF07 FF17 HH01 HH02 HH63
JJ24 LL01 LL02 TT01 TT02
5K101 NN08 SS06 SS08 TT06